FILE 'BIOSIS, MEDLINE, EMBASE, EMBAL, SCISEARCH, BIOTECHDS, CAPLUS'
ENTERED AT 18:15:45 ON 21 JAN 2003
L1 1 S BORK/AU
E BORK/AU
E BORK D/AU
L2 0 S E3 AND PREDICT?
L3 44463 S PREDICT? AND FUNCTION AND PROTEIN?
L4 3749 S L3 AND (COMPUT?)
L5 130 S L4 AND (ERROR? OR ERRON?)
L6 75 DUP REM L5 (55 DUPLICATES REMOVED)

L6 ANSWER 45 OF 75 MEDLINE
ACCESSION NUMBER: 1998196757 MEDLINE
DOCUMENT NUMBER: 98196757 PubMed ID: 9537411
TITLE: Predicting functions from protein
sequences--where are the bottlenecks?.
AUTHOR: Bork P; Koonin E V
CORPORATE SOURCE: EMBL, Heidelberg, Germany.. bork@embl-heidelberg.de
SOURCE: NATURE GENETICS, (1998 Apr) 18 (4) 313-8. Ref: 74
Journal code: 9216904. ISSN: 1061-4036.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199804
ENTRY DATE: Entered STN: 19980430
Last Updated on STN: 19980430
Entered Medline: 19980423
TI Predicting functions from protein sequences--where are
the bottlenecks?.
AB . . . of sequence data does not necessarily lead to an increase in
knowledge about the functions of genes and their products.
Prediction of function using comparative sequence
analysis is extremely powerful but, if not performed appropriately, may
also lead to the creation and propagation of assignment errors.
While current homology detection methods can cope with the data flow, the
identification, verification and annotation of functional features need.
. .
CT Check Tags: Animal; Human
Amino Acid Sequence
Computational Biology: MT, methods
Computational Biology: ST, standards
Databases, Factual
Molecular Sequence Data

*Proteins: GE, genetics
*Proteins: PH, physiology
Sequence Alignment: MT, methods
Sequence Alignment: ST, standards
Sequence Homology, Amino Acid
CN 0 (Proteins)

L6 ANSWER 29 OF 75 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V.
ACCESSION NUMBER: 2000203433 EMBASE
TITLE: Homology-based gene structure prediction: Simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps.
AUTHOR: Gotoh O.
CORPORATE SOURCE: O. Gotoh, Saitama Cancer Center, Research Institute, 818 Komuro Ina-machi, Saitama 362-0806, Japan. gotoh@cancer-c.pref.saitama.jp
SOURCE: Bioinformatics, (2000) 16/3 (190-202).
Refs: 47
ISSN: 1367-4803 CODEN: BOINFP
COUNTRY: United Kingdom
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 004 Microbiology
027 Biophysics, Bioengineering and Medical Instrumentation
LANGUAGE: English
SUMMARY LANGUAGE: English
TI Homology-based gene structure prediction: Simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps.
AB Motivation: Locating protein-coding exons (CDSs) on a eukaryotic genomic DNA sequence is the initial and an essential step in predicting the functions of the genes embedded in that part of the genome. Accurate prediction of CDSs may be achieved by directly matching the DNA sequence with a known protein sequence or profile of a homologous family member(s). Results: A new convention for encoding a DNA sequence into a series. . . this type of analysis. Using this convention, a dynamic programming algorithm was developed to align a DNA sequence and a protein sequence or profile so that the spliced and translated sequence optimally matches the reference the same as the standard protein sequence alignment allowing for long gaps. The objective function also takes account of frameshift errors, coding potentials, and translational initiation, termination and splicing signals. This method was tested on Caenorhabditis elegans genes of known structures. The accuracy of prediction measured in terms of a correlation coefficient (CC) was about 95% at the nucleotide level for the 288 genes tested,. . . and closest homologue